

# Dialogues

May 2019

## Statistics: Insight not numbers

by **Neil Sheldon**

*In this article, Neil Sheldon, chair of the Teaching Statistics Trust, description of article ...*

The past few years have seen big changes in the way that statistics is taught and learned in schools, particularly in the 16-19 age range. In England, for example, statistics has become – for the first time – a compulsory part of A-level mathematics. And many other subjects, such as biology, geography and psychology, now include some elements of statistics.

Traditionally, statistics at school level focused on process: routine calculations, plugging numbers into formulae and drawing graphs. But it is now increasingly being recognised that such tasks are best done by computers, not people. Indeed, in the age of ‘big data’ it is *only* computers that have the capacity to process the numbers. Modern approaches to statistics emphasise the importance of formulating problems and questions, interpreting results, carrying out inference and communicating conclusions. These activities have to be carried out by people, not computers – and the medium in which they are carried out is language.

It is important to realise that all of this goes deeper than just school examinations and qualifications. Statistical understanding is a key skill, one that everyone should have in order to make informed decisions. Indeed, it is not too grand a claim to say that statistical understanding is a democratic imperative if citizens are to play a full part in society. But statistical information in everyday life is often expressed in muddled, even incoherent, language – and so it is poorly understood. Furthermore, statistical inference, which requires careful and nuanced language to be done properly, is often absent or completely misconceived.

Students of statistics – and that should mean everyone – need to understand the meanings of terms (semantics) and the proper structure of sentences in which they are used (grammar).

Consider a simple example, the average: possibly the most common of all statistical ideas. The word ‘average’ can denote the arithmetic mean, the median (with 50% of the data above and 50% below) or the mode (the most common value). ‘Average’ can also mean ‘typical’ or ‘ordinary’. And, a point very often overlooked, the very concept of an average implies that there is variation from the average. There are linguistic issues aplenty lurking here.

A recent publication (*The Great British Family Report*, 2017, from the Nationwide Building Society) included the following statements.

- (a) The average family consists of two children
- (b) 27 is the average age to start a family
- (c) The average home has 3 bedrooms, 2 toilets, a family room and a toy room

Is statement (a) talking about the modal (most common) family size, or is it the mean rounded to a whole number? What type of average is 27 in statement (b), and is it the average for men, or for women, or for both parents combined? And what sort of average could it possibly be in statement (c) if the average family home has a toy room? The answers to those questions do not lie in numbers, but in language: specifically, language that is more precise and much clearer.

Semantic precision is necessary, but not sufficient. Even when it is clear what is meant by an average in a particular context, the surrounding language in which the word is embedded carries implications. This is a matter of getting the grammar right.

An internet search suggests that about 6% of the world's camels are Bactrians (with two humps), the rest being dromedaries (with one hump). So consider, from a linguistic point of view, the following sentences.

- (d) The average camel has 1.06 humps
- (e) Camels have, on average, 1.06 humps
- (f) The average number of humps per camel is 1.06

The topic of sentence (d) is 'the average camel'. The sentence carries some level of existential import. It assumes or implies or suggests that 'the average camel' exists – or at least that it is a useful concept. In fact, this is a very poor way of conveying the information. It is little more than a rather feeble joke.

Sentence (e) has 'camels' as its topic. It assumes that it makes sense to treat camels as a set of comparable objects. It takes *camel* to be a natural kind. In other words, its very construction obscures the distinction being made.

Sentence (f) has the topic 'humps'. The construction assumes that it makes sense to distribute humps among camels. It is like a latter-day *Just So* story: in giving out a hump to each camel, the Djinn finds that he has some left over – enough to give some of them a second hump.

The structures of these three sentences are faulty in that they make absurd assumptions. But such structures and assumptions are extremely common. And in circumstances where the context is a little more important than camels and their humps, there is serious danger of conceptual muddle and misunderstanding. The problem, again, does *not* lie in the numbers. It lies in the language. Analysing the assumptions inherent in the grammatical constructions is the key to understanding the issues.

It would be only fair at this point to ask what would be a good way of presenting the information about camels and humps. The best solution is to avoid talking about averages at all in this situation: most camels have one hump, with just 6% having two humps. Talking about the average can often be a lazy substitute for clarity of thought and expression.

The semantic and grammatical issues raised by the seemingly simple concept of an average are paralleled in the more complex area of statistical significance. The language of significance is widespread, but distinctions between everyday meanings and technical use are often not made and perhaps not often correctly understood.

In everyday use, 'significant' means something like 'being worthy of attention' or 'of practical importance'. The word acquired a technical meaning in connection with testing statistical hypotheses through the work of R. A. Fisher in the 1920s. Significance, in this sense, is a property of a set of data in relation to a hypothesis. It is a signal that the hypothesis might be worth further scrutiny in case it's not true. Unfortunately, it can be very difficult to know which of these meanings is intended. The following statements are taken from [an Ofqual publication](#) on examination grades.

- (g) The probability of receiving the definitive grade is significantly influenced by the overall spread of the grade boundaries.
- (h) The probability of achieving the definitive grade is not significantly different between the original and revised methods.
- (i) The two calculations do give rise to different statistics but, on the whole, these statistics are not significantly different.

It seems likely that, in statement (g), there is no technical meaning intended: the word 'significantly' could be replaced by 'considerably'. In (h), the matter is much less clear; this could be the report of a statistical test. In (i), it seems likely that there is no technical meaning, though the fact that it is two statistics which are being compared makes the interpretation more difficult. In each case, more careful use of language would be extremely helpful.

As with the average, it is not just the semantics that can mislead; there are grammatical considerations too. It violates the grammar of statistical inference to say things like 'that coin is significantly biased towards heads', or to talk about the probability of a hypothesis being true, as in 'given so many heads that coin is unlikely to be fair'. But this can be a difficult issue linguistically. Consider two similar questions.

- (j) What is the probability of a fair coin giving so many heads?
- (k) What is the probability it's a fair coin given so many heads?

The similarity here is only superficial, and these two questions are fundamentally different. Expressing them in more formal language, we get the following.

- (j') What is the probability of so many heads occurring given that the coin is fair?
- (k') What is the probability that the coin is fair given that so many heads have occurred?

These formulations make clear, through the main clauses, which is the event of interest (the number of heads in (j'), the coin being fair in (k')). The subordinate clauses ('given that ...') make clear what information is to be factored into the calculation.

Furthermore, it should now be clear that (k'), and hence (k), is an improper question. Statistics can tell us the probability of observations given certain assumptions. It cannot tell us the probability of the assumptions given the observations.

Pronouncing an observation to be statistically significant is a substitute for thought. It gives an illusion of certainty in an uncertain world.

If the new way of teaching and learning statistics is to be effective, teachers need to recognise that statistics is a discursive subject, be comfortable using and analysing language, be flexible and creative in their use of language when explaining and interpreting, and be rigorous in their use of language in order to achieve clarity and accuracy. And so do students.

The purpose of statistics is insight, not numbers, and the bridge from numbers to insight is language.

## Resources

Advice on writing about statistics, from the Writing Center at the University of North Carolina Chapel Hill: <https://writingcenter.unc.edu/tips-and-tools/statistics/>

Glossary provided by the Royal Statistical Society: <https://www.statslife.org.uk/resources/for-the-general-public/glossary>

## Further reading

Wasserstein, R. L. & Lazar, N. A. (2016). The ASA's statement on  $p$ -values: Context, process, and purpose. *The American Statistician*, 70(2), 129–133.

## About the author

Neil Sheldon is chair of the Teaching Statistics Trust and a former vice-president of the Royal Statistical Society. He is also vice-chair of the UK Linguistics Olympiad.